LIMIT DEXTRINASE INHIBITOR PROMOTER

1. FIELD OF INVENTION

The invention relates generally to a DNA sequence identified from *Hordeum vulgare* which acts as a tissue specific promoter. The invention further relates to the use of said DNA sequence to regulate exogenous gene expression in the endosperm and aleurone tissues of the seed of a plant.

2. BACKGROUND

The expression of a gene is dependent upon its DNA sequence being transcribed into RNA by the action of RNA polymerase. To achieve this, the RNA polymerase must recognise and attach to a region of DNA sequence located upstream of (i.e. 5' to) the gene coding sequence in order for transcription to be initiated. Such a region is termed the promoter of the gene. The intrinsic nature of the promoter sequence determines the circumstances and the manner in which the gene is expressed.

There are, broadly speaking, four types of promoters found in plant tissues; constitutive, tissue-specific, developmentally-regulated, and inducible/repressible.

A constitutive promoter directs the expression of a gene throughout the various parts of a plant continuously during plant development, although the gene may not be expressed at the same level in all cell types. Examples of known constitutive promoters include those associated with the cauliflower mosaic virus 35S transcript (Odell et al, 1985, Nature **313** 810-812), the rice actin 1 gene (Zhang et al, 1991, Plant Cell **3** 1155-1165) and the maize ubiquitin 1 gene (Cornejo et al, 1993, Plant Molec. Biol. **29** 637-646).

A tissue-specific promoter is one which directs the expression of a gene in one (or a few) parts of a plant, usually throughout the lifetime of those plant parts. The category of tissue-specific promoter commonly also includes tissue-preferred or tissue-enhanced promoters which may also direct expression at some level in tissues other than the preferred tissue. Examples of tissue-specific promoters known in the art include those associated with the patatin gene expressed in potato tuber and the high molecular weight glutenin gene expressed in wheat, barley or maize endosperm.

A developmentally-regulated promoter directs a change in the expression of a gene in one or more parts of a plant at a specific time during plant development. The gene may be

expressed in that plant part at other times during plant development at a different (usually lower) level, and may also be expressed in other plant parts.

An inducible promoter is capable of directing the expression of a gene in response to an inducer. In the absence of the inducer the gene will not be expressed. The inducer may act directly upon the promoter sequence, or may act by counteracting the effect of a repressor molecule. The inducer may be a chemical agent such as a metabolite, a protein, a growth regulator, or a toxic element, a physiological stress such as heat, wounding, or osmotic pressure, or an indirect consequence of the action of a pathogen or pest. A developmentally-regulated promoter might be described as a specific type of inducible promoter responding to an endogenous inducer produced by the plant or to an environmental stimulus at a particular point in the life cycle of the plant. Examples of known inducible promoters include those associated with wound response, such as described by Warner et al (1993), temperature response as disclosed by Benfey & Chua (1989), and chemically induced, as described by Gatz (1995).

A promoter sequence may comprise of a number of defined domains necessary for its function. The first of these defined domains comprises approximately 70 base pairs located immediately upstream of (5' to) the structural gene and forms the core promoter. The core promoter contains the CAAT and TATA boxes and defines the transcription initiation site for the structural gene. A series of regulatory sequences upstream of the core promoter constitute the remainder of the promoter sequence and determine the expression levels, the spatial and temporal patterns of expression, and the response to inducers. In addition some promoters contain sequence elements which act to enhance the level of expression, for example that from the pea plastocyanin promoter as described in International Patent Application, Publication No. WO 97/20056.

Genetic modification of plants depends upon the introduction of chimaeric genes into plant cells and their controlled expression under the direction of a promoter. Promoters may be obtained from different sources including plants, fungi, bacteria and viruses, and different promoters may work with different efficiencies in different tissues.

It may often be desirable to express introduced genes in a number of different tissues within a plant. For example, the expression of a resistance to a pathogen or pest, or tolerance to temperature extremes might be best expressed throughout all tissues in the plant. Similarly it might be desirable to ensure the expression of the transgenes at all times throughout the development of the plant. Also, a promoter which is expressed in a manner that is immune to the influence of inducers or repressors resulting from unforeseen environmental stimuli may

also be useful to ensure the continued expression of a trait. For these purposes, the use of a "constitutive" promoter would be desirable. Examples of constitutive promoters include the CaMv 35S promoter. For cereals the ubiquitin promoter is a constitutive promoter of choice (Christensen & Quail,1996)

However, in some instances it is more desirable to control the location of gene expression in a transgenic plant to enhance the effect of gene expression by ensuring that expression occurs preferentially in those tissues where the effect of the gene product is most efficacious. By the same argument, modulated expression can reduce potential yield loss by reducing the resource drain on the plant. Further advantages include limiting the expression of agronomically useful yet generally deleterious genes to specific tissues by localisation and compartmentalisation of gene expression in cases where the gene product must be restricted to, or excluded from, certain tissues. For example, anther specific expression of the *suc* inhibitor genes (Mariani et al, 1990, Nature 347 737) has been used in male sterility systems, whereas expression in other parts of the plant would result in toxicity.

Examples of endosperm specific barley promoters include the high molecular weight glutenin promoter (Halford *et al*). However, there is a lack of suitable endosperm specific promoters.

3. SUMMARY OF THE INVENTION

The present invention is directed to a DNA sequence identified from *Hordeum vulgare* (barley).

The present invention provides a DNA sequence comprising a promoter sequence, the DNA sequence comprising the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, said sequence being capable of regulating the expression of a gene.

The present invention also provides a recombinant DNA sequence, wherein said sequence comprises vector DNA and a DNA sequence being known herein as SEQ. ID. No.:1, or a portion thereof, or a sequence having at least 70% identity thereto, said sequence being capable of regulating the expression of a gene.

Suitably vector DNA may be, for example, DNA from a plasmid, cosmid or phage. Vectors will frequently include one or more selectable markers to enable selection of cells that have been transfected or transformed and to enable the selection of cells harbouring vectors

incorporating heterlogous DNA. Alternatively the selectable marker gene may be in a different vector to be used simultaneously with a vector containing the DNA of interest.

Examples of suitable marker genes include antibiotic resistance genes such as those conferring resistance to kanamycin, G418 and hygromycin (*npt-II, hyg*-B); herbicide resistance genes such as those conferring resistance to phosphinothricin and sulphonamide based herbicides (*bar* and *sul* respectively; EP-A-242246, EP-A-0369637) and screenable markers such as beta-glucuronidase (GB2197653), luciferase and green fluorescent protein.

The recombinant DNA may suitably further comprise the DNA coding sequence of a gene.

Preferably the recombinant DNA resides in a host cell suitable for transcription and translation.The present invention also provides a method of regulating the expression of a gene, the method comprising introducing into a plant a DNA sequence operably associated with the coding sequence of a gene, wherein said DNA sequence comprises the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, wherein said DNA sequence is capable of regulating the expression of a gene.

The present invention also provides a transgenic plant, the cells of which plant comprise a DNA sequence operably associated with a gene coding sequence, wherein said DNA sequence comprises the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, said DNA sequence being capable of regulating the expression of a gene..

The present invention further provides a method of modifying the metabolism within the cells of a transgenic plant the method comprising introducing into a plant a DNA sequence operably associated with the coding sequence of a gene, wherein said DNA sequence comprises the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, said DNA sequence being capable of regulating the expression of a gene.

The present invention even further provides a method of producing a gene product within the cells of a transgenic plant the method comprising introducing into a plant a DNA sequence is operably associated with the coding sequence of a gene, wherein said DNA sequence comprises the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, said DNA sequence being capable of regulating the expression of a gene.

The present invention also provides an oligonucleotide probe which selectively hybridizes to a DNA sequence, the DNA sequence comprising the sequence being known herein as SEQ. ID. No:1, or a portion thereof, or a sequence having at least 70% identity thereto, said DNA sequence being capable of regulating the expression of a gene.

Other suitable DNA sequences forming part of the present invention are DNA sequences that have at least 75%, 80%, 85%, 90% or 95% identity with the sequences listed herein and which retain the capability of regulating the expression of a gene.

Further suitable DNA sequences forming part of the present invention are DNA sequences which are complementary to the sequences listed herein.

Preferably the DNA sequence is capable of regulating the expression of a gene in the endosperm or aleurone tissues of developing seeds.

Preferably the DNA sequence is capable of regulating expression of a gene which encodes a limit dextrinase inhibitor protein (LDI).

More preferably the DNA sequence regulates expression of the gene encoding the Limit dextrinase inhibitor-like protein in *Hordeum vulgare.*

Most preferably the DNA sequence regulates expression of the gene encoding the Limit dextrinase inhibitor protein in the endosperm or aleurone tissues of developing seeds of *Hordeum vulgare.*

The DNA sequences referred to herein may be isolated DNA sequences or, alternatively, may be synthesised DNA sequences.

The present invention is applicable to all plants having a seeds with an endosperm and/or aleurone tissue.

In the present application the DNA sequence capable of regulating the expression of a gene may be a DNA sequence comprising a promoter sequence, the DNA sequence being selected from the group consisting of

a) a DNA sequence, said sequence being known herein as SEQ. ID. No:1;
b) a DNA sequence comprising a portion of the sequence of the isolated DNA of a) and being capable of regulating the expression of a gene;
c) a DNA sequence which is at least 95%, 90%, 85%, 80%, 75% or 70% homologous to the DNA sequence of a) or b);
d) a DNA sequence which is at least 80% identical to the DNA sequence of a) or b) and which is capable of regulating the expression of a gene;
e) a DNA sequence comprising a portion of the sequences of the DNAs of any one of a)-d) and being capable of regulating the expression of a gene; and

f) a DNA sequence which is complementary to the DNA sequence of any one of a) – e).

3.1 SEQUENCE IDENTIFIERS

In the sequence listing:

SEQ ID No. 1 shows the DNA sequence of the isolated barley limit dextrinase inhibitor promoter.

SEQ ID No. 2 shows the nucleotide and derived amino acid sequence of the isolated barley limit dextrinase inhibitor protein.

SEQ ID No. 3 shows the derived amino acid sequence of the isolated barley limit dextrinase inhibitor protein.

SEQ ID No. 4 shows the nucleotide sequence of the PCR primer inhib5.

SEQ ID No. 5 shows the nucleotide sequence of the PCR primer inhib6.

SEQ ID No. 6 shows the nucleotide sequence of the PCR primer AP1.

SEQ ID No. 7 shows the nucleotide sequence of the PCR primer GSP1.

SEQ ID No. 8 shows the nucleotide sequence of the PCR primer AP2

SEQ ID No. 9 shows the nucleotide sequence of the PCR primer GW1.

4. BRIEF DESCRIPTION OF THE FIGURES

In order that the invention may be readily carried into effect reference will now be made, by way of example, to the following diagrammatic drawings in which;

Figure 1 shows the results of an RT-PCR analysis of barley limit dextrinase inhibitor expression. The RT-PCR analysis of LDI expression was performed using Inhib 5 and Inhib 6 primers. Legend: λ/Hind III molecular weight markers (λ); complete barley grains 20 dpa (1); complete barley grains 40 dpa (2); endosperm of barley grains 14 dpa (3); aleurone layers of barley grains treated with gibberellic acid (4); negative control of RT-PCR (5); days post anthesis (dpa);

Figure 2 shows a the results of Northern blot analysis of limit dextrinase expression. The top panel of the blot shows the RNA gel and the bottom panel of the blot shows the Northern blot. Northern blot analysis of LDI expression was performed using LDI RNA probe on different RNA samples. Legend: Leaves in the light (LL); leaves in the dark (LD); roots (r); post anthesis (p.a.);

Figure 3 shows the results of Southern blot analysis of limit dextrinase inhibitor in barley genomic DNA. The Southern analysis of barley ("GoldenPromise") DNA was performed using limit dextrinase inhibitor DNA as a probe. Legend: λ/Hind III molecular weight markers (λ); DNA digested with EcoR V (E); DNA digested with BamH1 (B); DNA digested with Xho I (X);

Figure 4 shows an isolation of limit dextrinase inhibitor promoter fragments by genome walking PCR. DNA was digested with DraI, SmaI, EcoR V, Pvu II and Sca I. The primers used were AP2 and GW 1, and the PCR products analyzed on a preparative 1% TAE agarose gel. Legend: λ/Hind III molecular weight markers (λ); negative control of PCR reaction (-ve). Base pairs (bp);

Figure 5 shows a Barley transformation construct containing the limit dextrinase inhibitor promoter pCAMBIA1302 with LDI promoter;

Figure 6 shows a transient expression assay using the limit dextrinase promoter: GFP construct. Fluorescent microscopy was performed on endosperm 2-4 weeks post anthesis transiently assayed with an LDI Promoter: GFP construct. Legend: Non-bombarded controls (A, C, E and G); bombarded with LDI Promoter:GFP construct and analyzed for GFP expression 24h and 48h after bombardment (B, D, F and H). Scaling bar is 100μm in length.

Figure 7 shows the putative transcription start sites and TATA sequences (-25 bp) within SEQ. ID. No: 1 as predicted by neural network (M.G. Reese, N.L. Harris & F.H. Eeckman. 1996; Large Scale Sequencing Specific Neural Network for promoter & slice site recognition: Biocomputing Proceedings of 1996 Pacific Symposium). The putative transcript start site is shown in larger font.

Figure 8 shows the results of a full sequence search using the BLASTN algorithm for sequences similar to SEQ. ID. No. 1 of the present invention. The search results provided no significant homology with any known sequence in the four databases searched.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to plant gene regulatory sequences. Specifically this invention relates to a promoter identified in *Hordeum vulgare* (barley).

A plant DNA sequence may be recovered from the cells of the natural host or it may be synthesized directly *in vitro*. Extraction from the natural host enables the isolation *de novo* of novel sequences, whereas *in vitro* DNA synthesis generally requires pre-existing sequence information. Direct chemical *in vitro* synthesis can be achieved by sequential manual synthesis or by automated procedures. DNA sequences may also be constructed by standard

techniques of annealing and ligating fragments or by other methods known in the art. Examples of such cloning procedures are given in Sambrook et al.

The DNA sequence of the present invention may be isolated by direct cloning of segments of plant genomic DNA and screening the DNA sequence for the presence of diagnostic sequence motifs characteristic of known promoter sequences. Suitable segments of genomic DNA may be obtained by fragmentation using restriction endonucleases, sonication, physical shearing, or other methods known in the art.

The identification of the cloned segment as a promoter sequence may alternatively be achieved by assessing functionality, for example by linking the cloned segment with a coding sequence derived from a reporter gene and introducing the chimaeric construct into a host cell or cell-free system wherein expression of the reporter gene can be evaluated. This process may form part of another sequence isolation strategy termed promoter trapping, wherein genomic DNA fragments are cloned directly into "expression vectors" comprising a reporter gene coding region and other sequences necessary for expression in a host cell or cell-free system. The expression may or may not require integration of the chimaeric construct into the host's chromosomal DNA (Topping J F. *et al.*(1991) and Topping J F. *et al.* (1994)).

An alternative method of obtaining DNA sequences of the present invention is by the identification and isolation of a DNA coding sequence which is known to be expressed and subsequently using this sequence to obtain the contiguous promoter sequence, which is by definition directing the expression of the coding sequence. A coding sequence may be obtained by the isolation of messenger RNA (mRNA or polyA+ RNA) from plant tissue. The tissue used for RNA isolation is selected on the basis that suitable gene coding sequences are believed to be expressed in that tissue at optimal levels for isolation.

Various methods for isolating mRNA from plant tissue are well known to those skilled in the art, including for example using an oligo-dT oligonucleotide immobilised on an inert matrix. The isolated mRNA may be used to produce its complementary DNA sequence (cDNA) by use of the enzyme reverse transcriptase (RT) or other enzymes having reverse trancriptase activity. Isolation of an individual cDNA sequence from a pool of cDNAs may be achieved by cloning into bacterial or viral vectors, or by employing the polymerase chain reaction (PCR) with selected oligonucleotide primers. The production and isolation of a specific cDNA from mRNA may be achieved by a combination of the reverse transcription and PCR steps in a process known as RT-PCR.

In order to isolate the promoter sequence of the present invention, a cDNA sequence to a Limit dextrinase inhibitor protein may be identified. Examples of other Limit dextrinase

inhibitor protein genes are given in Genbank Accession No.s X13443 for the *Hordeum vulgare* alpha-amylase/trypsin inhibitor; X99982. for the *Triticum aestivum* cDNA for the PUP88 protein ; AJ222975 for the *Hordeum sponteneum* cDNA for the Itr1 gene for BTI-Cme2.2 protein; AP005197 for the *Oryza sativa* cDNA for the putative hageman factor inhibitor protein ; X61032 for the *Triticum durum* cDNA for the alpha amylase inhibitor protein; and X54064 for the *Zea mays* cDNA for the Hageman factor inhibitor protein

Limit dextrinase inhibitor proteins suitable for use in isolating the promoter of the present invention include proteins having the amino acid sequence given as SEQ. ID. No. 3 and proteins homologous to; and having essentially the same biological properties as, the protein disclosed herein as SEQ. ID. No. 3. This definition is intended to encompass natural allelic variations in the protein plus any non-allelic examples. It will be appreciated that the amino acid sequence need not be identical to that of SEQ. ID. No. 3; for the purposes of this invention, the amino acid sequence may be at least 80%, 90%, or 95% homologous or more with the protein of SEQ. ID. No:3 to retain its biological activity. General categories of potentially equivalent amino acids include, but are not limited to: glutamic acid and aspartic acid; lysine, arginine and histidine; alanine, valine, leucine and isoleucine; asparagine and glutamine; threonine and serine; phenylalanine, tyrosine and tryptophan; and glycine and alanine.

The selected cDNA may then be used to evaluate the genomic features of its gene of origin by use as a hybridisation probe in a Southern blot of plant genomic DNA to reveal the complexity of the genome with respect to that sequence. Alternatively, sequence information from the cDNA may be used to devise oligonucleotides and these may be used in the same way as hybridisation probes for PCR primers to produce hybridisation probes or for PCR primers to be used in direct genome analysis.

Similarly the selected cDNA may be used to evaluate the expression profile of its gene of origin by use as a hybridisation probe in a Northern blot of RNA extracted from various plant tissues, or from a developmental or temporal series. Again sequence information from the cDNA may be used to devise oligonucleotides which can be used as hybridisation probes to produce hybridisation probes or directly for RT-PCR.

The selected cDNA, or derived oligonucleotides, may then be used as a hybridisation probe to challenge a library of cloned genomic DNA fragments and identify overlapping DNA sequences. By this means a contiguous promoter may be identified and isolated.

By the nature of the method of isolation, an isolated cDNA usually comprises the 3' terminus of the coding region and extends towards the 5' terminus. It may not comprise the

full-length coding sequence. It is preferable to ensure that the 5' terminal sequence is present if the cDNA is to be used to identify the contiguous promoter. This may be achieved by extension of the cloned cDNA sequence in the 5' direction by a process termed 5' RACE.

If sequence analysis of the cloned cDNA identifies a homologous sequence already reported in the scientific literature, this information may provide a suitable candidate sequence for the 5' terminus. However the possibility of there being different members of the same gene family with similar coding regions, but differing intron regions, promoter sequences and expression profiles may lead to the selection of an incorrect and unsuitable promoter sequence.

Once the 5' terminus of the coding sequence has been identified, the contiguous upstream region containing the promoter may be isolated by further extension in the 5' direction. This may be achieved by methods including vector-ligation PCR, genome walking, vectorette PCR, and other methods. If necessary the process may be repeated with a new primer complementary to the 5' terminus of the first promoter fragment to ensure that all the control sequences of the promoters are isolated.

The present invention provides a DNA sequence which controls the expression of a Limit dextrinase inhibitor protein which DNA sequence is known herein as SEQ. ID. No. 1, and which may be used as a plant promoter. The present invention also includes within its scope DNA sequences which control the expression of other Limit dextrinase inhibitor protein genes and which are homologous with the DNA sequence of SEQ. ID. No. 1.

Homology may be determined on the basis of percentage identity between two DNA (or polypeptide) sequences. "Percentage identity", as known in the art, is a measure of the relationship between two polynucleotides or two polypeptides, as determined by comparing their sequences. In general the two sequences to be compared are aligned to give a maximum correlation between the sequences. The alignment of the two sequences is examined and the number of positions giving an exact nucleotide (or amino acid) correspondence between the two sequences determined, divided by the total length of the alignment multiplied by 100 to give a percentage identity figure. This percentage identity figure may be determined over the whole length of the sequences to be compared, which is particularly suitable for sequences of the same or very similar lengths and which are highly homologous, or over shorter defined lengths, which is more suitable for sequences of unequal length or which have a lower level of homology.

For example, sequences can be aligned with the software clustalw under Unix which generates a file with the ".aln" extension, this file can then be imported into the Bioedit program (Hall, T.A. 1999. Bioedit: a user-friendly biological sequence alignment editor and

analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41:95-98) which opens the .aln file. In the Bioedit window, one can choose individual sequences (two at a time) and align them. This method allows for comparison of entire sequences.

Methods for comparing the identity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J. et al, Nucleic Acids Res. **12** 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the percentage identity between two polynucleotides and the percentage identity between two polypeptide sequences. BESTFIT uses the "local homology" algorithm of Smith and Waterman (Advances in Applied Mathematics, **2** 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences which are dissimilar in length, the program assuming that the shorter sequence represents a portion of the longer. In comparison, GAP aligns two sequences finding a "maximum similarity" according to the algorithm of Needleman and Wunsch (J. Molec. Biol. **48** 443-454, 1970). GAP is more suited to comparing sequences which are approximately the same length and an alignment is expected over the entire length. Preferably the parameters "Gap Weight" and "Length Weight" used in each program are 50 and 3 for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, percentage identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Karlin & Altschul, 1990, *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin & Altschul, 1993, *Proc. Natl. Acad. Sci. USA* 90:5873-5877, available from the National Center for Biotechnology Information (NCB), Bethesda, Maryland, USA and accessible through the home page of the NCBI at www.ncbi.nlm.nih.gov). These programs exemplify a preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al., 1990, *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., 1997, *Nucleic Acids Res.* 25:3389-3402.

Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, 1988, *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

Another non-limiting example of a program for determining identity and/or similarity between sequences known in the art is FASTA (Pearson W.R. and Lipman D.J., Proc. Nat. Acac. Sci., USA, 85:2444-2448, 1988, available as part of the Wisconsin Sequence Analysis Package). Preferably the BLOSUM62 amino acid substitution matrix (Henikoff S. and Henikoff J.G., Proc. Nat. Acad. Sci., USA, 89:10915-10919, 1992) is used in polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Yet another non-limiting example of a program known in the art for determining identity and/or similarity between amino acid sequences is SeqWeb Software (a web-based interface to the GCG Wisconsin Package: Gap program) which is utilized with the default algorithm and parameter settings of the program: blosum62, gap weight 8, length weight 2.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

Preferably the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a polynucleotide or a polypeptide sequence of the present invention, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value.

In the context of the present invention "substantially homologous" sequences are those which have at least 45% sequence identity, preferably at least 50% sequence identity, more preferably at least 60%, 70%, 80%, or 90% sequence identity and most preferably at least 95% sequence identity with the described sequences. In some cases sequence identity may be 98% or more preferably 99%, or more.

The present invention also includes DNA sequences which hybridize to the above isolated DNA, including partial sequences and complementary sequences. Conditions under

which such sequences will so hybridize can be determined in a routine manner. For example, the present invention includes within its scope, DNA sequences which hybridise to the DNA sequences of the present invention under stringent conditions. In the context of the present invention, "stringent conditions" are defined as those given in Martin *et al* (EMBO J 4:1625-1630 (1985)) and Davies *et al* (Methods in Molecular Biology Vol 28: Protocols for nucleic acid analysis by non-radioactive probes, Isaac, P.G. (ed) pp 9-15, Humana Press Inc., Totowa N.J, USA)). The conditions under which hybridization and/or washing can be carried out can range from 42°C to 68°C and the washing buffer can comprise from 0.1 x SSC, 0.5 % SDS to 6 x SSC, 0.5 % SDS. Typically, hybridization can be carried out overnight at 65°C (high stringency conditions), 60°C (medium stringency conditions), or 55°C (low stringency conditions). The filters can be washed for 2 x 15 minutes with 0.1 x SSC, 0.5 % SDS at 65°C (high stringency washing). The filters were washed for 2 x 15 minutes with 0.1 x SSC, 0.5 % SDS at 63°C (medium stringency washing). For low stringency washing, the filters were washed at 60°C for 2 x 15 minutes at 2 x SSC, 0.5% SDS.

The present invention also includes DNA sequences which hybridize to oligonucleotide probes. Preferably the DNA sequences hybridize to oligonucleotide probes under stringent conditions. In instances wherein the nucleic acid molecules are oligonucleotides ("oligos"), highly stringent conditions may refer, for example, to washing in 6x SSC / 0.05% sodium pyrophosphate at 37°C (for 14 base oligos), 48°C (for 17 base oligos), 55°C (for 20 base oligos), and 60°C (for 23 base oligos).

The gene coding sequence employed in carrying out the present invention may be active in some or all plant tissues. The sequence employed may encode a protein or an RNA moiety. Through recombinant DNA techniques the sequence may encode a synthetic variant of a protein or RNA, a partial sequence, or a composite sequence comprising regions from one or more genes. For example, the chimaeric gene may encode a polyprotein. The sequence may also comprise repeated, inverted or complementary sequences, to achieve disruption of the transcription and translation of an endogenous gene or genes by, for example, antisense or RNA inhibition technology.

The sequence known herein as SEQ. ID. No. 2 includes a coding sequence therein for the gene encoding a limit dextrinase inhibitor protein. In addition, many plant, bacterial and viral genes may be actively expressed. Exemplary genes encode the GUS, GFP and luciferase enzymes which may be used as reporter genes for promoter function. In addition the LDI promoter may be used to direct the expression of DNA sequences designed to alter the

metabolism of the endosperm of the seeds of cereal plants, especially barley. Such genes include those for carbohydrate metabolism, starch metabolism, amino acid and protein metabolism and lipid metabolism.

The gene product of the present invention can be produced by recombinant techniques, wherein genomic DNA clones or cDNA clones for the DNA coding sequence are produced, isolated, proliferated, and incorporated into a plant transformation vector of the present invention.

Plant transformation vectors

Plant transformation vectors of the present invention will contain "expression cassettes" comprising 5'-3' in the direction of transcription, a promoter sequence as described in the present invention, a gene coding sequence as discussed above and, optionally a 3' untranslated, terminator sequence including a stop signal for RNA polymerase and a polyadenylation signal for polyadenylase.

The promoter sequence may be present in one or more copies, and such copies may be identical or variants of the promoter sequence as described above. Such copies may also be complete or partial sequences as described above.

The terminator sequence may be obtained from plant, bacterial or viral genes. Suitable terminator sequences are the pea *rbcS* E9 terminator sequence, the *nos* terminator sequence derived from the nopaline synthase gene of *Agrobacterium tumefaciens* and the 35S terminator sequence from cauliflower mosaic virus, for example. A person skilled in the art will be readily aware of other suitable terminator sequences.

The expression cassette may also comprise a gene expression enhancing mechanism to increase the strength of the promoter. An example of such an enhancer element is that derived from a portion of the promoter of the pea plastocyanin gene, and which is the subject of International Patent Application, Publication No. WO 97/20056.

These regulatory regions may be derived from the same gene as the promoter DNA sequence of the present invention or may be derived from different genes, e.g. from *Hordeum vulgare* or other organisms. All of the regulatory regions should be capable of operating in cells of the tissue to be transformed.

The gene coding sequence may be derived from the same gene as the promoter DNA sequence of the present invention or may be derived from a different gene, e.g. from *Hordeum vulgare* or another organism.

The expression cassette may be incorporated into a basic plant transformation vector, such as, binary or super-binary vectors, for example, pBIN19 and pBIN+ and other suitable plant transformation vectors known in the art.

In addition to the expression cassette, the plant transformation vector will contain such sequences as are necessary for the transformation process. These may include the *Agrobacterium vir* genes, one or more T-DNA border sequences, and a selectable marker or other means of identifying transgenic plant cells.

Production of the gene product

Expression of a DNA coding sequence in the plant host cell will produce an RNA transcript. If the coding sequence is derived from a structural gene, the RNA transcript is then translated into a protein gene product. If desired, the gene product can be isolated by standard techniques for isolating proteins from biological systems, such as salt precipitation, column chromatography, immunoaffinity techniques, electrophoresis, recrystallisation, centrifugation, and such like.

Example 1. Isolation of barley *limit dextrinase inhibitor* cDNA fragments.

Cloning of LDI gene by RT-PCR

Total RNA was extracted from barley (var. Golden Promise) grains 2 and 4 weeks post anthesis with a LiCl method as described by (Cathala *et al.*, 1983).

2 μg of RNA was treated with Rnase-free DNase I (Amersham Pharmacia Biotech) and used to synthesize 30 μl first strand cDNA using random hexamer primers (Roche) and M-MLV reverse transcriptase (Promega) using the reaction conditions recommended by the manufacturer.

It will be recognised by one skilled in the art that other mRNA extraction and cDNA synthesis methods exist which could be employed to produce cDNA from tissue of *Hordeum vulgare.*

A 3 μl aliquot of the cDNA product was used in a standard PCR reaction containing 2 mM $MgCl_2$, 8% (v/v) DMSO and primers Inhib-5 (5'-ACCAATAAACTAGTATCAACAATGGCATCCGACCA-3' SEQ ID No 4) and Inhib-6 (5'-CCAACCTTTTTTATTCATCAATCGGCCACA-3' SEQ ID No 5), which were designed against the *Hordeum vulgare* Limit dextrinase inhibitor sequence (Genbank Accession no. X13443.), using Taq polymerase (Bioline) as recommended by the manufacturer. The PCR program used was 5 min at 94 °C, followed by 35 cycles of 94 °C for 30 sec, 63.5 °C for 30

sec and 72 °C for 1 min, finalized by 7 min at 72 °C. The product length was 623 bp. The amplified product was cloned into a pGEM-T Easy vector (Promega) and verified by sequencing.

It will be recognised by one skilled in the art that other methods exist which could be employed to produce cloned DNA fragments.

The sequence of this cDNA clone is shown in SEQ ID No 2. Compared to the published sequence (Genbank Accession no. X13443.), there are five single base pair substitutions, which in turn lead to two amino acid substitutions.

Example 2. RT-PCR analysis of Limit Dextrinase Inhibitor expression.

Total RNA was isolated from barley tissues using a LiCl method (M.-J.Cho *et al.*, 1998). 2 µg of total DNAse treated RNA was used to produce cDNA with M-MLV reverse transcriptase. 1/10 of this cDNA was used to run a PCR reaction with specific primers for the entire coding region of LDI.

The expression of the LDI gene was investigated by using RT-PCR with the specific primers Inhib 5 (SEQ ID No. 4) and Inhib 6 (SEQ ID No. 5) on RNA from developing grains. Expression of LDI was found in grains 20 and 40 days post anthesis, the endosperm and gibberellic acid treated aleurone layers. The highest level of expression was found in the endosperm. As shown in Figure 1.

Example 3. Northern analysis of Limit Dextrinase Inhibitor expression.

Total RNA was isolated from developing and germinating barley grains, barley leaves in the light or dark, and barley roots using a LiCl method (M.-J.Cho et al., 1998). 5 µg of total RNA were separated on a 1 % agarose gel with formaldehyde. The RNA was blotted over night with 10xSSC onto a positively charged Nylon membrane and UV crosslinked. The blot was then prehybridized in DIG-EasyHyb and hybridised with 100 ng/mlDIG-labeled RNA probe at 68 °C over night. The blot was washed once in 2xSSC, followed by 2 washes with 0.2xSSC at 68 °C. After that the blot was blocked for 1 hour at RT in blocking reagent. This was followed by 30 minutes in 1:10,000 dilution of anti-DIG antibody solution. To remove excess antibody the blot was washed 4x in Tris buffer with salt, then equilibrated in phosphatase buffer and the chemiluminiscent substrate (CDP-star) added. Chemiluminescence detection using X-ray film was performed. Northern blotting and hybridisation was carried out as detailed by manufacturers of the DIG label (Roche).

Expression of LDI was only detected in early development of barley grains (2 and 4 weeks post anthesis). No expression was found in germinating grains and vegetative tissue as shown in Figure 2.

Example 4. Southern blotting

Genomic DNA of barley leaf tissue was isolated according to the method of A.W.MacGregor *et al.*, 1995 and S.L.Delaporta *et al.*, 1983. 10 µg digested genomic DNA (isolated using the Ambion Phytopure Kit according to manufacturer's instructions) was separated on a 0.8 % TAE agarose gel at 20V over night. The DNA was depurinated in 250 mM HCl for 10 minutes, denatured for 30 min in 0.4 M NaOH and neutralized before blotting for 8 hours onto a positively charged Nylon membrane (Roche) in 20x SSC. The DNA was UV crosslinked and prehybridized in DIG-Easyhyb for 1 hour at 52 °C and hybridised with 40ng/ml DIG labeled DNA probe at 52 °C over night. Washing and detection was carried out as described before for Northern blotting except for a more stringent wash with 0.1 x SSC.

Southern blot analysis was used to examine the number of gene copies in wildtype (wt) "Golden Promise" barley. The exact number of copies of LDI in wildtype "Golden Promise" barley is hard to estimate as several bands per digest are visible. The most likely explanation is that there is a small gene family of related genes present in barley, as shown in Figure 3.

Example 5. Isolation of the limit dextrinase inhibitor promoter region by genome walking.

Genomic DNA of barley leaf tissue was isolated according to the method of A.W.MacGregor etal., 1995 and S.L.Delaporta et al., 1983. 2.5 µg genomic barley DNA (variety "Golden Promise") was digested with the restriction enzymes Sca I, Pvu II, EcoR V, Sma I or Dra I respectively. These DNA fragments were then ligated to an asymmetric adapter and used as a template for PCR reactions with adapter primers and gene specific primers to isolate the 5' region of the LDI gene. All procedures were carried out as described in the Clontech manual. The first PCR was carried out using primers AP1 (SEQ ID No. 6) and GSP1 (SEQ ID No. 7). 1µl of this 50 µl PCR reaction was used as template for the secondary (nested) PCR using primers AP2 (SEQ ID No. 8) and GW1 (SEQ ID No. 9).

Using the procedures described above, two candidate fragments were isolated, from digests with Sma I and Pvu II, as shown in Figure 4. The DNA fragments from the Sma I (~1000 bp) and Pvu II (~ 650 bp) digests were purified and cloned into the Topo 4 vector and sequenced. Both fragments (Sma I and Pvu II) showed the same A and T rich sequence,

spanning the 790 bp long 5' region of LDI. The promoter sequence is shown in SEQ ID No. 1. SEQ. ID. No. 1 discloses a nucleic acid sequence of 833 bp, within which sequence there is the 790 bp promoter sequence of the present invention. The sequence also contains therewithin the 5' untranslated region of the gene coding sequence, up to and including the first codon of the gene.

Example 6. Transient Expression of Limit Dextrinase Inhibitor Promoter marker gene constructs.

In order to investigate if the 790 bp region upstream of the LDI gene was sufficient as a promoter, a vector containing the putative promoter region plus a reporter gene was constructed. It consists of the 790 bp *Sma* I fragment, the GFP gene (green fluorescent protein) and a terminator (NOS). This construct (Figure 5) was used for transient assays on barley endosperm (2-4 weeks post anthesis) and for stable barley transformation. Endosperms were transformed with a biolistic device using .construct DNA coated gold particles and analysed by fluorescence microscopy after 24 and 48 hours post bombardment. The results in shown in Figure 6 demonstrate the expression of GFP, showing that the isolated genomic sequence acts as promoter.

Example 7. Stable expression of Limit Dextrinase Inhibitor Promoter marker gene constructs.

Stable barley transformation of immature embryos of the barley variety "Golden Promise" was performed with a biolistic device using DNA coated gold particles. Selection and regeneration procedures were performed as described by Y.Wan & P.G.Lemaux, 1994.

References.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25** 3389-3402

Ausubel, F.M., Bent, R., Kingston, R.E., Moore, D.D., Seidmann, J.G., Smith, J.A. & Struhl, K. (1989). *Current protocols in molecular biology* . New York: Greene and Willy Interscience.

Benfey, P.N., and Chua, N-H. (1989) Regulated genes in transgenic plants. Science **244**, 174-181.

Bevan, M. (1984) Binary *Agrobacterium* Vectors for Plant Transformation. Nucleic Acids Research **12** 8711-8721

Buluwela, L., Forster, A., Boehm, T. & Rabbitts, T.H. (1989). A rapid procedure for colony screening using nylon filters. *Nucleic Acids Research* **17**, 452-452.

Christensen, A. H. & Quail, P.H. (1996). Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Research* **5**, 213-218

Cho, M.-J., Jiang, W. & Lemaux, P.G. (1998). Transformation of recalcitrant barley cultivars through improvement of regenerability and decreased albinism. *Plant Science* **138**, 229-244.

Conkling, M.A., Cheng, C-L., Yamamoto, Y.T. and Goodman, H.M. (1990) Plant Physiol. **93** 1203-1211

Delaporta, S.L., Wood, J. R. & Hicks, J.B. (1983). A plant DNA minipreparation: version 2.*Plant Molecular Biology Reporter* **1**, 19-22.

Gatz, C. (1995) Novel inducible/repressible gene expression systems. Methods in Cell Biol. **50**, 411-424

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eichholtz, D., Rogers, S.G. & Fraley, R.T. (1985) A simple and general method for transferring genes into plants. Science **227**, 1229-1231.

Jefferson, R.A. (1987) Assaying Chimeric Genes in Plants: The GUS Gene Fusion System. Plant Molec. Biol. Reptr. **5** 387-405

Kristensen, M., Planchot, V., Abe, J.-I. & Svensson, B (1998). Large-scale purification and characterization of barley limit dextrinase, a member of the a-amylase structural family. *Cereal Chemistry* **75**, 473-479.

Lazaro, A., Rodriguez-Palenzuela, P., Marana, C., Carbonero, P. & Garciaolmedo, F. (1988). Signal peptide homology between the sweet protein thaumatin II and unrelated cereal alpha-amylase/trypsin inhibitors. *FEB* **239**, 147-150.

Lemaux, P.G. Cho, M.-J., Zhang, S. & Bregitzer, P. (1999). Transgenic cereals: Hordeum vulgare L. (barley). In *Molecular improvement of cereal crops*, ed. I.K.VASIL, pp. 255-316. Dordrecht: Kluwer Academic Publishers.

Macgregor, A.W., Macri, L.J., Bazin, S.L. & Sadler, G.W. (1995). Limit dextrinase inhibitor in barley and malt and its possible role in malting and brewing. *EBC Congress* 185-193.

Macgregor, A.W., Macri, L.J., Schroeder, S.W. & Bazin, S.L. (1994). Purification and characterisation of limit dextrinase inhibitors from barley. *Journal of Cereal Science* **20**, 33-41.

Macgregor, E.A., Bazin, S.L., Ens, E.W., Lahnstein, J., Macri, L.J., Shirley N.J. & Macgregor, A.W. (2000). Structural models of limit dextrinase inhibitors from barley. *Journal of Cereal Science* **31**, 79-90.

Manners, D.J. (1997). Observations on the possible involvement of debranching enzymes in starch biosynthesis. *J.Appl.Glycosci.* **44**, 541-546.

Sambrook, J., Frisch, E.F. and Maniatis, T. (Eds.) (1989) Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Topping J F; Wei W; Lindsey K; (1991) Development 112, 1009-1019. Functional tagging of regulatory elements in the plant genome

Topping J F; Agyeman F; Henricot B; Lindsey K (1994) The Plant Journal 5, 895-903, Identification of molecular markers of embryogenesis in Arabidopsis thaliana by promoter trapping.)

Wan, Y. & Lemaux, P.G. (1994). Generation of large numbers of independently transformed fertile barley plants. *Plant Physiology* **104**, 37-48.